Partie II

Suite du cours

"Analyse de Survie"

3 - Estimateur de la fct de risque h :

On a : $h(t) = \frac{f(t)}{S(t)}$

un estimateur de h : aux points t_i (réalisations de T_i).

$$\boxed{\hat{h}_n(t_i) = \frac{m_i}{n_i}}$$

où m_i = nbre d'ev^t observés à t_i

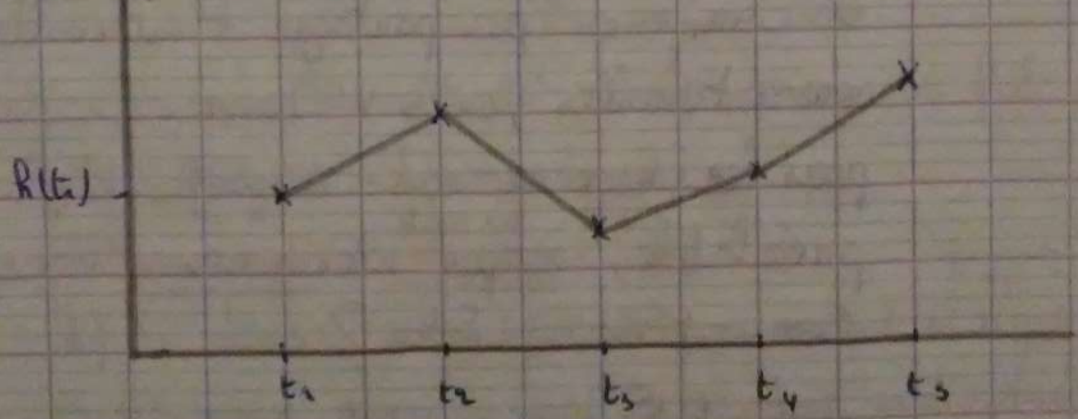
n_i = nbre d'individus à risque à t_i^- (avant t_i).

c'est le rapport entre le nbre d'év^t observés et les cas possibles

<u>Rappel</u> : $h(t) := \lim_{\Delta \to 0} \frac{\mathbb{P}(t < T < t+\Delta \mid T > t)}{\Delta}$

on fait une interpolation linéaire entre les points $(t_i, h(t_i))$

(ou polynômes, splines ...)



<u>Rq</u> : Il existe d'autres méthodes pour estimer la fonction de risque h.

§. Estimation de la survie S par la méthode acturielle (Actuariat)

L'Actuariat est une science qui s'interesse aux contrats d'assurance contre les risques.

Dans l'estimation de S par Kaplan-Meier, on a les instants t_i de la survenue de l'evenement E ou censure, par contre pour assurer (compagnie de l'assurance) un individu contre des risques (voitures, habitation ...) on ne dispose pas des points t_i à l'avance et on voudrait une estimation de S, ($S(t)$ = la "survie" à l'instant t) il existe une methode appelée Methode Actuarialle ou Actuariat

§ 2. Methode actuarielle :

on commence par partager l'intervalle $[0, +\infty[$ en sous-intervalles fixés : $]0, a_1]$, $]a_1, a_2]$, ..., $]a_{j-1}, a_j]$

pour le choix de sous-intervalles il y a une etude préalable, et ne sont pas nécessairement égaux.

Comment estimer $S(t)$?

Soit $t \in]a_{j-1}, a_j]$ $j = 1, \dots$:

$$S(t) = \mathbb{P}(T > t) = \mathbb{P}(T > t / T > a_{j-1})\mathbb{P}(T > a_{j-1})$$

alors :

$$S(a_j) = \underbrace{\mathbb{P}(T > a_j / T > a_{j-1})}_{= q_j} \mathbb{P}(T > a_{j-1})$$

Ainsi : $S(a_j) = q_j \, q_{j-1} \cdots q_1$
$= \prod_{i=1}^{j} q_i$

où $q_i = \mathbb{P}(T > a_i / T > a_{i-1})$
$=$ proba de survie ds $]a_{i-1}, a_i]$
on a pas l'év^t ds cet intervalle

q_j = proba de ne pas avoir E ds $]t_{j-1}, t_j]$
$p_i = 1 - q_i$ = proba d'avoir E ds $]a_{i-1}, a_i]$
Pour estimer q_i on estime p_i
On note : e_i = le nbre d'individus à risque ds l'intervalle $]a_{i-1}, a_i]$.
m_{i-1} = nbre d'individus qui ont subi l'évenement ds $]a_{i-1}, a_i]$.
alors $\boxed{p_i \approx \dfrac{m_{i-1}}{e_i} =: \hat{p}_i}$

La détermination des e_i et des m_{i-1} se fait par les compagnie d'assurance

* Détermination des e_i

1) Si dans l'intervalle $(a_{i-1}, a_i]$ il n'y a pas de censure, avec n_i = nbre d'individus à risque à l'instant a_i
on a : $n_i = n_{i-1} - m_{i-1}$.
alors on choisit (un choix !) : $\boxed{e_i = n_{i-1} = n_i + m_{i-1}}$

2) s'il y a des données censurées ds $]a_{i-1}, a_i]$ (la censure ici correspond à des individus qui se présentent (observés) au début puis disparaissent par la suite)

on choisit (choix!) $\boxed{e_i = n_{i-1} - \frac{c_{i-1}}{2}}$

où c_{i-1} = nbre d'individus censurés, ds $]a_{i-1}, a_i]$.

n_{i-1} avant a_{i-1}

Donc l'estimateur actuariel de S :

$\hat{S}_a(a_j) = \prod_{i=1}^{j} \hat{q}_i$ où $\hat{q}_i = 1 - \frac{m_{i-1}}{e_i}$

et $e_i = n_{i-1} - \frac{c_{i-1}}{2}$

$$\boxed{\hat{S}_a(a_j) = \prod_{i=1}^{j} \left(1 - \frac{m_{i-1}}{n_{i-1} - \frac{c_{i-1}}{2}}\right)} \quad j \geq 0$$

Reste $\hat{S}_a(t)$: on fait une interpolation linéaire des points $(a_j, \hat{S}_a(a_j))$ (car $\hat{S}_a$ n'est cste ds l'intervalle $]a_{j-1}, a_j]$ car il y a des évènements à l'intérieur.

• Estimateur de Nelson Aalen

La fonction de risque cumulée :

$H(t) = \int_0^t h(s)\,ds$ $\quad (\leftrightarrow F(t) = \int_0^t f(s)\,ds)$

(somme des risques (risque cumulé))

Estimation de H :

On a : $H'(t) = h(t)$.

or : $H(t_i + \Delta) - H(t_i) = \int_{t_i}^{t_i+\Delta} h(s)\,ds \simeq \Delta . h(t)$.

ou encore

$$H(t_i) - H(t_i^-) \simeq h(t_i).$$

or $h(t_i) \simeq \frac{m_i}{n_i}$

donc un estimateur de $H(t_i)$

$$\hat{H}_n(t_i) - \hat{H}_n(t_i^-) = \frac{m_i}{n_i}$$

d'où $\boxed{\hat{H}_{Aa}(t) = \sum_{t_i \leq t} \frac{m_i}{n_i}}$ est de **Nelson Aalen**

car : $\hat{H}_n(t_i) = \sum_{t_j \leq t_i} \frac{m_j}{n_j}$

$$\hat{H}_n(t_i^-) = \sum_{t_j \leq t_i^-} \frac{m_j}{n_j}$$

$H(t)$ c'est la somme des risques jusqu'à t_i

$H(t_i^-)$ " " " t_i^-

En faisant la différence il reste le risque à l'instant t_i.

D'autres estimateurs de H :

De la formule $H(t) = -\log S(t)$,

~~[illegible]~~ on a d'autres estimateurs de H :

$\hat{H}_{1,KM}(t) = -\log \hat{S}_{KM}(t)$, $\hat{H}_a(t) = -\log \hat{S}_a(t)$

§ Propriétés des estimateurs de S

L'étude des propriétés asymptotiques des estimateurs $\hat{S}_{KM}$, $\hat{S}_a$, $\hat{H}_{Nn}$, ... est liée au processus ponctuels

$$\begin{cases} \text{martingales et différence des martingales,} \\ N_t = M_t + A_t \end{cases}$$

On arrive à :

$E(\hat{\theta}) = \theta$

$E(\hat{S}_{KM}(t)) \underset{n \to +\infty}{\longrightarrow} S(t)$ (estimateur sans biais asympt)

e.s.b

$E(\hat{S}_a(t)) \to S(t)$

$S_{KM}(t) \to S(t)$ en proba

conv. en loi.

. Estimation des quantiles :

. Médiane : $F(M) = \frac{1}{2}$. $S(M) = \frac{1}{2}$

Si dans une étude de survie on vient de tracer $S(\hat{S}_{KM}, \hat{S}_a, \dots)$

Comment estimer la médiane M :

$$\boxed{\hat{M}_n \text{ vérifie : } \hat{S}_{KM}(\hat{M}_n) = \frac{1}{2}}$$

Estimation des ~~estim~~ quantiles : α - quantiles : q_α

estimer $\hat{q}_\alpha$ par : $\boxed{\hat{S}_{KM}(\hat{q}_{\alpha,n}) = \alpha}$

On montre que $\hat{q}_{\alpha,n} \to q_\alpha$.

$\cancel{(\hat{q}_{\alpha,n} \to q_\alpha)}$

$S_{KM}(\hat{M}) \to S(M) = \frac{1}{2}$